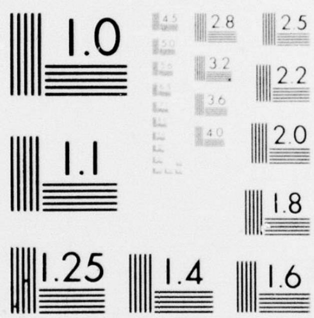
AD-A037 764 CALIFORNIA UNIV LOS ANGELES DEPT OF SYSTEM SCIENCE F/G 12/1
LIKELIHOOD RATIOS FOR TIME-CONTINUOUS DATA MODELS: THE WHITE NO--ETC(U)
DEC 76 A V BALAKRISHNAN AF-AFOSR-2492-73

UNCLASSIFIED AFOSR-TR-77-0173 NL

1 OF 1
AD
A037764

END

DATE
FILMED
4-77



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ADA037764

19 REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 18 AFOSR - TR - 77 - 0173

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): 6 LIKELIHOOD RATIOS FOR TIME-CONTINUOUS DATA MODELS: THE WHITE NOISE APPROACH

5. TYPE OF REPORT & PERIOD COVERED: 9 Interim rept.

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): 10 A. V. Balakrishnan

8. CONTRACT OR GRANT NUMBER(s): 15 AF-AFOSR [redacted]-2492-73

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
University of California, Los Angeles
System Science Department
Los Angeles, CA 90024

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 17 16 2304/A1

11. CONTROLLING OFFICE NAME AND ADDRESS:
Air Force Office of Scientific Research (NM)
Building 410
Bolling AFB, Washington, D.C. 20332

12. REPORT DATE: 11 Dec[redacted]76

13. NUMBER OF PAGES: 10

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): 12 11p.

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DDC
APR 4 1977
C

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

COPY AVAILABLE TO DDC DOES NOT PERMIT FULLY LEGIBLE PRODUCTION

18. SUPPLEMENTARY NOTES

IRIA International Symposium on New Trends in Systems Analysis, Versailles, France, December 1976, 9 pages, Dec 1976.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

BEST AVAILABLE COPY

DDC FILE COPY

This document

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

~~We~~ develops a formula for likelihood functionals for signals in additive noise in the time-continuous case using a white noise approach. It is shown that the formula differs from the well-known formula in the Wiener process version by the appearance of an additional term corresponding to the conditional mean square filtering error.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED 405 213

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

AFOSR - TR - 77 - 0173

Likelihood Ratios for Time-Continuous Data Models: The White Noise Approach

A. V. Balakrishnan†
System Science Department
University of California

Paper presented at the IRIA International Symposium on New Trends in Systems Analysis, December 1976, Versailles, France.

†Research supported in part under Grant No. 73-2492, Applied Mathematics Divn. AFOSR, USAF

Likelihood Ratios for Time-Continuous Data Models: The White Noise Approach

A. V. Balakrishnan
System Science Department
University of California

Abstract We develop a formula for likelihood functionals for signals in additive noise in the time-continuous case using a white noise approach. It is shown that the formula differs from the well-known formula in the Wiener process version by the appearance of an additional term corresponding to the conditional mean square filtering error.

1. Introduction. In much of engineering literature on identification (too voluminous to be referred to individually. See the several volumes of proceedings of IFAC Symposia on System Identification and Parameter Estimation, 1970, 1973 and 1976) it is customary to consider the observed data as sampled periodically in time -- even when the basic phenomena are modelled by time-continuous differential equations. The usual 'hand-waving' argument is then made that the 'limiting' continuous-time case is no more than a mathematical detail; and that anyhow in digital computer processing, conversion to sampled data is a basic step. This is indeed true; but the authors almost invariably proceed to use the model:

$$y_n = S_n + N_n$$

where $\{S_n\}$ is the information-bearing time series and $\{N_n\}$ the observation noise series, and (this is the crucial point) take $\{N_n\}$ as a sequence of independent variables. But this requires that the sampling rate be not more than twice the noise bandwidth, itself unknown. Of course, to answer this objection, one can allow $\{N_n\}$ to be correlated; but then the correlation function must be known. Now it is ver unrealistic to require the correlation function of instrument noise; and even when known, it adds a lot to the complication but little to the performance. We maintain that it is much better to use a time-continuous model

$$y(t) = S(t) + N(t) \tag{2}$$

and allow the sampling rate to be as high as the A-D converter wants to use. But in the time-continuous model we are faced with another problem; the basic tool in identification is the likelihood ratio (for fixed parameters): the Radon-Nikodym derivative of the probability measure induced by the process $y(\cdot)$ to that induced by $N(\cdot)$. But this likelihood ratio is difficult to implement even when the spectrum of $N(\cdot)$ is known, which it is not. What we can say for sure is that the bandwidth of the (instrument) noise is large compared with that of the process $S(\cdot)$. At this point the earlier engineering literature used the notion of "white noise" a process with constant spectral density over all frequencies in a formal way. In the sixties it became fashionable to replace this by the "Wiener process" model as "more rigorous". Thus we replace (2) by

$$Y(t) = \int_0^t S(\sigma)d\sigma + W(t) \tag{3}$$

where $W(t)$ is a Wiener process. We have then, to be sure, the advantage of the powerful machinery of Martingales and Ito integrals. In fact the likelihood functional (for the case where signal and noise are independent which we assume thruout) can then be written down as: [see [1]]:

$$\mathrm{Exp} - 1/2\left\{\int_0^T ||\hat{S}(t)||^2 dt - 2\int_0^T [\hat{S}(t), dY(t)]\right\} \tag{4}$$

where $\hat{S}(t)$ is the best mean square estimate of $S(t)$ given the observation $Y(s)$ upto time t. But the hooker is that the second term is an Ito integral:

$$\int_0^T [\hat{S}(t), dY(t)]$$

This integral is defined on the basis that $Y(t)$ is of unbounded variation with probability one. On the other hand no physical instrument can produce such a wave form! Indeed, we must now go back to (2) where it came from and thus replace

$$dY(t) \text{ by } y(t)dt$$

This is all right if $S(t)$ is deterministic; if not, we no longer obtain the value prescribed by the Ito formula! In particular, any algorithm based on it leads to erroneous results. Most authors of papers on the subject probably have never bothered to calculate anything based on actual data; and of course in any digital computer simulation it is possible to mask this completely. Indeed, almost all simulation models employ the discrete version (1).

Faced with this difficulty we have to examine more precisely the model again. Thus what we want to exploit is the fact that the bandwidth of the noise is large compared to that of the process $S(\cdot)$. Hence what is really needed is the 'asymptotic form' of the likelihood functional as the bandwidth goes to infinity in an arbitrary manner.

Such a theory has been developed by the author using a precise notion of white noise. See [2] for details. We take the 'sample points' to be in a Hilbert Space with Gauss measure theorem. Thus in (2) we consider $N(t)$ $0 < t < T$ as pathwise square integrable in $[0,T]$; as elements in the L_2-space $L_2(R_n; (0,T))$, (the observation having its range in R_n, n-dimensional Euclidean Space). Corresponding to white noise with 'unit' spectral density, we define the Gauss measure by:

$$E[e^{i\int_0^T [N(t),\, h(t)]dt}] = \text{Exp} - 1/2 \int_0^T [h(t),\, h(t)]dt$$

for each $h(\cdot)$ in $L_2[R_n; (0,T)]$, defining thus the characteristic function of the Gauss measure.

The difference between this set-up and the Wiener-process set-up is simply this. Let $\{\phi_n(\cdot)\}$ denote a complete orthornormal system in $L_2[R_n, (0,T)]$. Then

$$\int_0^T [\phi_n(t), N(t)]dt = \zeta_n$$

yield a sequence of zero-mean, unit variance Gaussians. The sample-space for the sequence is ℓ_2, since

$$\sum_1^\infty \zeta_n^2 = \int_0^T N(t)^2 dt < \infty$$

On the other hand, given such a sequence it is standard practice to take R^∞ as the sample space and via the Kolmogorov theory, construct a countably additive measure on the Borel sets of R^∞. [This is also the countably additive extension to Nuclear Spaces via the Minlos theorem]. This is in fact the Wiener process theory, in which of course, all of ℓ_2 has zero measure. Both set ups of course agree on the measures of cylinder sets. What is rendered difficult by using ℓ_2 as the sample space is the notion of a random variable. Whereas this is trivial in the R^∞ model -- any Borel-measurable function being a random variable -- it is the central issue in the ℓ_2 - set-up. In other words, given any functional $f(\cdot)$ on $L_2[R_n; (0,T)]$, even continuous thereon, it need not define a random variable. We define it as a random variable if and only if for any sequence P_n of finite dimensional projections converging strongly to the identity, the sequence $\{f(P_n(\cdot)\}$ is Cauchy in probability, and all such sequences are equivalent. Thus we have a smaller class of random variables; the implication being that the Ito integrals in the Wiener process theory may not correspond to random-variables on ℓ_2. Moreover the 'limiting' notion corresponds to the 'bandwidth expanding' notion.

2. Likelihood Ratio: White Noise Theory.

Let us now examine likelihood ratios (Radon Nikodym derivatives) in terms of the white noise theory. Let

$$y(t) = S(t) + N(t) \quad 0 \leq t \leq T < \infty \tag{2.1}$$

where $S(\cdot)$ and $N(\cdot)$ are independent processes. We shall assume that the signal $S(\cdot)$ has finite energy:

$$\int_0^T E(||S(t)||^2)dt < \infty \tag{2.2}$$

For each t, $0 < t \leq T$, let

$$W(t) = L_2[R_n; (0,t)]$$

We shall shorten $W(T)$ to simply W. Under condition (2.2), the process $S(\cdot)$ induces a countably additivie measure on W (and hence on W(t) for each t). [The cylinder measure on W can be extended to be countably additive, in other words; this is a consequence of the Sazonov theorem, see [2]]. Thus (2.1) defines a weak distribution on W defined by the characteristic function:

$$E[e^{i[y,h]}] = C_s(h) \ \ \mathrm{Exp} - 1/2 \ ||h||^2 \tag{2.3}$$

where

$$C_s(h) = E[e^{i[S,h]}] \tag{2.4}$$

where we have used the inner-product notation:

$$[S,h] = \int_0^T [S(t), h(t)]dt, \ h \in W.$$

Then the cylinder measure μ_y induced by $y(\cdot)$ is absolutely continuous with respect to Gauss measure μ_G and the Radon-Nikodym derivative is defined by the function:

$$f(\omega) = \int_W \mathrm{Exp} - 1/2 \ \{||S||^2 - 2\ [S,\omega)\} \ d\mu_s \tag{2.5}$$

Thus for any cylinder set C,

$$\mu_y(C) = \underset{n\to\infty}{\text{limit}} \int_C f(P_n\omega)\, d\mu_G$$

where P_n is any sequence of finite dimensional projections strongly convergent to the identity. This result has been proved in [3].

Let $\{\phi_n\}$ be an orthonormal basis in W and let L denote the mapping of W into ℓ_2:

$$Lx = a;\ a_n = \int_0^T [x(\sigma), \phi_n(\sigma)]d\sigma.$$

Let

$$LS = \zeta$$

Let μ_ζ denote the measure induced on ℓ_2 by this mapping. Then we can rewrite (2.5) in the form

$$f(\omega) = \int_{\ell_2} \text{Exp} - 1/2\ \{[\zeta,\zeta] - 2\,[\zeta, L\omega]\}\ d\mu_\zeta \tag{2.6}$$

It must be emphasised that (2.6) is defined for every element ω in W. Note also that (2.6) can be defined with respect to any orthonormal system $\{\phi_n\}$.

The likelihood functional f(y) where $y(\cdot)$ is the observation, will now be expressed in a form similar to (4). For this purpose, let (2.6) be defined with respect to the orthonormal system $\{\phi_n\}$. For each t, $0 < t \leq T$, define the operators $\Lambda(t)$, mapping W into ℓ_2 by:

$$\Lambda(t)x = a;\ a_n = \int_0^t [\phi_n(\sigma), x(\sigma)]d\sigma \tag{2.7}$$

Let

$$R(t) = \Lambda(t)\ \Lambda(t)^*. \tag{2.8}$$

Then the Radon-Nikodym derivative of the measure induced by the process $y(\cdot)$ over [0,t] with respect to Gauss measure on W(t) is given by:

$$f(t,\omega) = \int_{\ell} \operatorname{Exp} - 1/2 \{[R(t)\,\zeta,\zeta] - 2\,[\zeta, \Lambda(t)\omega]\}\, d\mu_\zeta \tag{2.9}$$

Note that $(T) = L$. Let P_n denote the projection operator corresponding to the first n basis functions $\{\phi_1\}$, $i = 1,\ldots n$. Then we define

$$\hat{\zeta}(t) = \lim_{n} E[\zeta | \Lambda(t)\, P_n y] \tag{2.10}$$

As shown in [3], we have (Bayes Formula) that

$$(t) = \frac{\int_{\ell_2} \zeta \operatorname{Exp} - 1/2 \{[R(t)\zeta,\zeta] - 2\,[\zeta, \Lambda(t)y]\}\, d\mu_\zeta}{\int_{\ell_2} \operatorname{Exp} - 1/2 \{[R(t)\zeta,\zeta] - 2\,[\zeta, \Lambda(t)y]\}\, d\mu_\zeta} \tag{2.11}$$

Note that, by Schwartz Inequality

$$||\hat{\zeta}(t)||^2 \le \frac{\int_{\ell_2} ||\zeta||^2 \operatorname{Exp} - 1/2 \{[R(t)\zeta,\zeta] - 2\,[\zeta, \Lambda(t)y]\}\, d\mu_\zeta}{\int_{\ell_2} \operatorname{Exp} - 1/2 \{[R(t)\zeta,\zeta] - 2\,[\zeta, \Lambda(t)y]\}\, d\mu_\zeta}$$

$$= \frac{\int_{\ell_2} ||\zeta||^2 \operatorname{Exp} - 1/2\, ||\, R(t)\zeta - \Lambda(t)y||^2\, d\mu_\zeta}{\int_{\ell_2} \operatorname{Exp} - 1/2\, ||\, R(t)\zeta - \Lambda(t)y||^2\, d\mu_\zeta}$$

$$\le c\, E[||\zeta||^2] \operatorname{Exp} + 1/2\, (||\, \Lambda(t)y|| + k)^2,\ 0 < c,\ k < \infty \tag{2.12}$$

It should be noted that such an estimate is not available in the Wiener process version. Moreover we shall show that (2.9) is actually absolutely continuous in t with an L_2-derivative. Let $\phi(t)$ be infinitely differentiable with compact support in $(0,T)$. Then

$$\int_0^T [f(t,\omega)\, \phi'(t)]dt$$

$$= \int_{\ell_2} \int_0^T [\operatorname{Exp} - 1/2 \{[R(t)\zeta,\zeta] - 2\,[\zeta, \Lambda(t)\omega]\}\, \phi'(t)dt]\}\, d\mu_\zeta$$

$$= \int_{\ell_2} \left(\int_0^T - 1/2 \, ||\sum_1^\infty \phi_i(t)\zeta_i||^2 + [\sum_1^\infty \phi_i(t)\zeta_i, \omega(t)] \right) \left(\text{Exp} - 1/2 \, \{[R(t)\zeta,\zeta] \right.$$

$$\left. - 2 \, [\zeta, \Lambda(t)\omega]\} \quad \phi(t)dt \right) d\mu_\zeta \qquad (2.13)$$

where we note that both

$$||\sum_1^\infty \phi_i(t)\zeta_i||^2 \quad \text{and} \quad [\sum_1^\infty \phi_i(t)\zeta_i, \omega(t)]$$

are in L_2 [0,T] for each ζ in ℓ_2. Hence the derivative is (defined a.e. $0 < t < T$):

$$\int_{\ell_2} (-\tfrac{1}{2}||\sum_1^\infty \phi_i(t)\zeta_i||^2 + [\sum_1^\infty \phi_i(t)\zeta_i, \omega(t)]) \ \text{Exp} - 1/2 \, \{[R(t)\zeta,\zeta] - 2 \, [\zeta, \Lambda(t)\omega]\} \, d\mu_\zeta$$

we shall next prove that

$$g_N(t) = \sum_1^N \phi_i(t)\hat{\zeta}_i(t) \qquad 0 \leq t \leq T$$

converges in the norm of W. But this is immediate from the fact that, analogous to (2.12):

$$||g_N(t)||^2 \leq E[||\sum_1^N \phi_i(t)\zeta_i||^2] \ \text{Exp} + 1/2 \, ||\Lambda(t)y||^2 \qquad \text{a.e.} \quad 0 < t < T$$

Let

$$\hat{S}(t) = \sum_1^\infty \phi_i(t)\hat{\zeta}_i(t)$$

and

$$\widehat{||S(t)||^2} = \frac{\int_{\ell_2} ||\sum_1^\infty \phi_i(t)\zeta_i||^2 \ \text{Exp} - 1/2 \, \{[R(t)\zeta,\zeta] - 2 \, [\zeta, \Lambda(t)y]\} \, d\mu_\zeta}{\int_{\ell_2} \text{Exp} - 1/2 \, \{[R(t)\zeta,\zeta] - 2 \, [\zeta, \Lambda(t)y]\} \, d\mu_\zeta}$$

Then from (2.13) we can write:

$$\frac{d}{dt} \ \text{Log} \ f(t, y) = - 1/2 \, \{||\hat{S}(t)||^2 - 2 \, [\hat{S}(t), y(t)] + \widehat{||S(t)||^2} - ||\hat{S}(t)||^2\}$$

and hence finally, for the log likelihood functional:

Log f(y)

$$= -1/2\left\{\int_0^T \|\hat{S}(t)\|^2 dt - 2\int_0^T [\hat{S}(t), y(t)]dt + \int_0^T [\widehat{\|S(t)\|^2} - \|\hat{S}(t)\|^2]\, dt\right\} \tag{2.13}$$

we note that the third term can also be expressed as

$$\lim_{n\to\infty} E[\|S(t) - \hat{S}(t)\|^2 \mid \Lambda(t)P_n y] \tag{2.14}$$

The formula (2.13) differs from the Wiener process version in the appearance of the third term; in the case where $S(t)$ is Gaussian, we know that (2.14) reduces to

$$E[\|S(t) - \hat{S}(t)\|^2]$$

which is then also independent of the observation $y(\cdot)$; see [3]. Note that (2.14) can be large in the case where the noise level is large. Formula (2.13) was derived in [3] for a seemingly less general case by a different method. Finally we remark that (2.13) is consistent with the 'circle differential' formalism of Ito [4].

References

1. Y. Liptser and A. Shiryayev: Statistics of Random Processes, Nanka 1974.
2. A. V. Balakrishnan: Applied Functional Analysis, Springer-Verlag, 1976.
3. A. V. Balakrishnan: Radon-Nikodym Derivatives of a class of weak distributions on Hilbert Spaces, Journal of Applied Mathematics and Optimization, 1977 (to be published).
4. K. Ito: Stochastic Differentials, Journal of Applied Mathematics and Optimization Vol. 1, No. 4, 1975.

FILM

4